Intro to Semantic Web

Presented by: Aidan Hogan
Slides by EUCLID and Aidan
WHAT IS THE SEMANTIC WEB?
... machine readable Web?
... benvolent machines?
... a bunch of standards?
... a cake?

Images by Hendler, Brickley Novack; [http://www.bnode.org/blog/tag/layer%20cake](http://www.bnode.org/blog/tag/layer%20cake)
2.2 Interpretations

Given a datatype map $D$ and a vocabulary $V$ over $D$, an interpretation $I = (\Delta_I, \Delta_D, C, OP, DP, I, DT, LT, FA)$ for $D$ and $V$ is a 9-tuple with the following structure:

- $\Delta_I$ is a nonempty set called the object domain.
- $\Delta_D$ is a nonempty set disjoint with $\Delta_I$ called the data domain such that $(DT)^{DT} \subseteq \Delta_D$ for each datatype $DT \in V_D$.
- $C$ is the class interpretation function that assigns to each class $C \in V_C$ a subset $(C)^C \subseteq \Delta_I$ such that
  - $(\text{owl:Thing})^C = \Delta_I$ and
  - $(\text{owl:Nothing})^C = \emptyset$.
- $OP$ is the object property interpretation function that assigns to each object property $OP \in V_{OP}$ a subset $(OP)^{OP} \subseteq \Delta_I \times \Delta_I$ such that
  - $(\text{owl:topObjectProperty})^{OP} = \Delta_I \times \Delta_I$ and
  - $(\text{owl:bottomObjectProperty})^{OP} = \emptyset$.
- $DP$ is the data property interpretation function that assigns to each data property $DP \in V_{DP}$ a subset $(DP)^{DP} \subseteq \Delta_I \times \Delta_D$ such that
  - $(\text{DataProperty})^{DP} = \Delta_I \times \Delta_D$ and
  - $(\text{objectProperty})^{DP} = \emptyset$.
- $\text{interpretation function that assigns to each individual } a \in V_I \text{ an element } (a)^I \in \Delta_I$.
- $\text{interpretation function that assigns to each datatype } DT \in V_D \text{ a subset } (DT)^{DT} \subseteq \Delta_D$ such that
  - as in $D$ for each datatype $DT \in N_D$.
- $\text{interpretation function defined as } (n)^{LT} = (LV, DT)^{LS}$ for each $n \in V_{LT}$, where $LV$ is the lexical form of $n$.
- $\text{interpretation function defined as } (F, lt)^{FA} = (F, (lt)^{LT})^{LS}$ for each $(F, lt) \in V_{FA}$.

Define the extensions of $OP$, $DT$, and $C$ to object property expressions, data ranges, and class expressions.

- $OP$ is extended to object property expressions as shown in Table 1.
... free/open data?

Image from http://www.digitaltimes.ie
... linking stuff?
... naming everything?
... Web 3.0?


Image by Radar Networks; Nova Spivak; http://memebox.com/futureblogger/show/824

03.10.2013
... a buzzword?!
… the research area your supervisor assigned you to but that you don’t really understand?
WHY IS THE SEMANTIC WEB?
Great Wave of Data

Wikipedia
≈ 5.9 TB of data
(Jan. 2010 Dump)

1 Wiki = 1 Wikipedia
Great Wave of Data

Human Genome

≈ 4 GB/person
≈ 0.0006 Wiki/person
Great Wave of Data

US Library of Congress
≈ 235 TB archived
≈ 40 Wiki
Great Wave of Data

Sloan Digital Sky Survey
≈ 200 GB/day
≈ 73 TB/year
≈ 12 Wiki/year
Great Wave of Data

NASA Center for Climate Simulation
≈ 32 PB archived
≈ 5,614 Wiki
Great Wave of Data

Facebook
≈ 12 TB/day added
≈ 2 Wiki/day
≈ 782 Wiki/year
(as of Mar. 2010)
Great Wave of Data

Large Hadron Collider
≈ 15 PB/year
≈ 2,542 Wikipedias/year
Great Wave of Data

Google
≈ 20 PB/day processed
≈ 3,389 Wiki/day
≈ 7,300,000 Wiki/year
(Jan. 2010)
Great Wave of Data

Internet (2016)
≈ 1.3 ZB/year
≈ 220,338,983 Wiki/year
(2016 IP traffic; Cisco est.)
Data Bottleneck

← Rate at which data are produced

← Rate at which data can be understood
WHAT PROBLEM DOES IT SOLVE?
... a personal answer

c.a. 2005
... the research area your supervisor assigned you to but that you don’t really understand?
The Dessert Problem
The Dessert Problem

• Requirements:
  – Must be a dessert
  – Must be citrus-free
  – Must be gluten-free
  – Ingredients available in local supermarket(s)
  – Cheap (students)
  – Must be delicious
The Dessert Problem
Dessert Algorithm: Naive

candidates := ∅

for all recipe-site in google-results
   for all dessert-recipe in recipe-site
      if dessert-recipe type looks-delicious
         suitable := true
         for all ingredient in dessert-recipe
            if searchNutrition(ingredient) type wheat or lemon or lime ... suitable := false
            else if searchShops(ingredient) type null or expensive suitable := false
            end
         end
         if suitable add dessert-recipe to candidates end
      end
   end
end

return candidates
Dessert Algorithm: Utopia

candidates := ∅

for all safe-cheap-local-dessert-recipe in magical-sem-web-results()
    if safe-cheap-local-dessert-recipe type looks-delicious
        add safe-cheap-local-dessert-recipe to candidates
    end
end

return candidates
The Dessert Solution?

magical-sem-web-results()
BEYOND DESSERT ...
Music!

• Provision of a music-based portal.
• Bring together a number of disparate components of data-oriented content:

1. **Musical content** (streaming data & downloads)
2. Music and artist metadata
3. Review content
4. **Visual content** (pictures of artists & albums)
Music!

Visualization Module

Analysis & Mining Module

RDFa

Metadata

SPARQL Endpoint

Integrated Dataset

LD Wrapper

D2R Transf.

LD Wrapper

RDF/XML

Streaming providers

Downloads

Metadata

Other content

LD Dataset

Access

Application

Vocabulary Mapping

Interlinking

Cleansing

Physical Wrapper

Musical Content

Data acquisition

EUCLID

Educational Curriculum for the usage of Linked Data
SEMANTIC WEB FOUNDATIONS
Internet

The growth of the Internet

Source: http://www.evolutionoftheweb.com
The Web

• There is a wealth of information on the Web.

• It is aimed mostly towards consumption by humans as end-users:
  • Recognize the meaning behind content and draw conclusions,
  • Infer new knowledge using context and
  • Understand background information.
The Web

• Billions of diverse documents online, but it is not easily possible to automatically:
  • Retrieve relevant documents.
  • Extract information.
  • Combine information in a meaningful way.

• Idea:
  • Also publish machine processible data on the web.
  • Formulate questions in terms understandable for a machine.
  • Do this in a standardized way so machines can interoperate.

• The Web becomes a **Web of Data**
  • This provides a common framework to share knowledge on the Web across application boundaries.
The Web: Evolution

Web of Documents → Web of Data

Hyperlinks → Typed Links

"Documents" → "Things"
The Web: Evolution

Source: http://www.radarnetworks.com
Web Technology Basics

HTML – HyperText Markup Language
• Language for displaying web pages and other information in a web browser.
• HTML elements consist of tags (enclosed in angle brackets), attributes and content.

HTTP – HyperText Transfer Protocol
• Foundation of data communication for the WWW.
• Client-server protocol.
• Every interaction is based on: request and response.
Uniform Resource Identifier (URI)

- Compact sequence of characters that identifies an abstract or physical resource.

- Examples:
  - ldap://[2001:db8::7]/c=GB?objectClass?one
  - mailto:John.Doe@example.com
  - news:comp.infosystems.www.servers.unix
  - tel:+1-816-555-1212
  - telnet://192.0.2.16:80/
  - http://dbpedia.org/resource/Karlsruhe
CORE OF THE SEMANTIC WEB
Semantics on the Web

Semantic Web Stack
Berners-Lee (2006)
Semantics on the Web

- **Syntatic basis**
  - Basic data model
  - Expressive vocabulary (ontology) language
  - Query language

- **Semantic Web Stack**
  - Simple vocabulary (schema) language
  - Expressive vocabulary (ontology) language
  - Query language

- **Application specific declarative-knowledge**

- **Query language**
  - SPARQL
  - OWL
  - RIF

- **Data interchange: RDF**

- **User Interface & applications**

- **Trust**
  - Digital signatures, recommendations
  - Proof generation, exchange, validation

- **Crypto**

- **Semantic Web Stack**
  - Berners-Lee (2006)
Semantics on the Web

RDF – Resource Description Framework

Semantic Web Stack
Berners-Lee (2006)
RDF – Resource Description Framework

- RDF is the basis layer of the Semantic Web stack ‘layer cake’.

- Basic building block: RDF triple.
  - **Subject** – a thing (identified by URI)
  - **Predicate** – a relationship (identified by URI)
  - **Object** – another thing (identified by URI) or a literal

- **Subject** has relationship **Predicate** to **Object**
- *(Tom_Scholz has relationship **lead_singer** to Boston)*
Why URIs?

RDF – Resource Description Framework

- **Subject** has relationship **Predicate** to **Object**
- *(Tom_Scholz has relationship **lead_singer** to Boston)*

- URIs:
  - Avoid ambiguity!
  - Enable linking (discussed later)
  - Enable dereferencing (discussed later)
Semantics on the Web

RDF – Resource Description Framework (Example)

URLs are given in angle brackets in N-Triples.

In N-Triples every statement is terminated with a full stop.

Literal are given in quotes in N-Triples.
Semantics on the Web

RDF Graphs

• Every set of RDF assertions can then be drawn and manipulated as a (directed labelled) graph:
  • **Resources** – the subjects and objects are nodes of the graph.
  • **Predicates** – each predicate use becomes a label for an arc, connecting the subject to the object.
Semantics on the Web

RDF Graphs (Example)

<http://musicbrainz.org/artist/b10bbbfc-cf9e-42e0-be17-e2c3e1d2600d#>

<http://www.w3.org/2002/07/owl#sameAs>

<http://dbpedia.org/resource/The_Beatles>

"The Beatles"

<http://xmlns.com/foaf/0.1/name>

"Subject"

"Predicate"

"Object"
RDF Graphs (Example)

<http://musicbrainz.org/artist/b10bbbc-cf9e-42e0-be17-e23e1d2600d#>

<http://www.w3.org/2002/07/owl#sameAs>

<http://xmlns.com/foaf/0.1/name>

<http://dbpedia.org/resource/The_Beatles>

"The Beatles"

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

<http://dbpedia.org/ontology/Music_Artist>
RDF Blank Nodes

• RDF graphs can also contain unidentified resources, called *blank nodes*:

```
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.w3.org/2003/01/geo/wgs84_pos#Point>
<http://www.w3.org/2003/01/geo/wgs84_pos#lat>
<http://www.w3.org/2003/01/geo/wgs84_pos#long>
```

• Blank nodes can group related information, but their use is often discouraged
RDF Turtle

- Turtle is a syntax for RDF more readable.

- Since many URIs share same basis we use prefixes:
  @prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
  @prefix rdfs:<http://www.w3.org/2000/01/rdf-schema#>.
  @prefix owl:<http://www.w3.org/2002/07/owl#>.
  @prefix mo:<http://purl.org/ontology/mo/>.
  @prefix dbpedia:<http://dbpedia.org/resource/>.

And (sometimes) a unique base:
@base <http://musicbrainz.org/>.
Semantics on the Web

RDF Turtle

• Also has a simple shorthand for class membership:

```
@base <http://musicbrainz.org/>.
@prefix mo:<http://purl.org/ontology/mo/>.
<artist/b10b3bfc-cf9e-42e0-be17-e2c3e1d2600d#> a mo:MusicGroup.
```

Is equivalent to:

```
'http://musicbrainz.org/artist/b10b3bfc-cf9e-42e0-be17-e2c3e1d2600d#'
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
```
Semantics on the Web

RDF Turtle

• When multiple statements apply to same subject they can be abbreviated as follows:

```rdfs:label "The Beatles"; owl:sameAs dbpedia:The_Beatles, <http://www.bbc.co.uk/music/artists/b10b
```
RDF Turtle

• Turtle also provides a simple syntax for **datatypes and language tags** for literals, respectively:

```turtle
<recording/5098d0a8-d3c3-424e-9367-1f2610724410#_> a mo:Signal;
   rdfs:label "All You Need Is Love";
   mo:duration "PT3M48S"^^xsd:duration .

dbpedia:The_Beatles dbpedia-owl:abstract

   "The Beatles were an English rock band formed (...) "@en,
   "The Beatles waren eine britische Rockband in den (...) "@de .
```
Semantics on the Web

RDF/XML

• Common misconception: RDF is not exotic XML!
  – RDF is triples!

• RDF/XML widespread, but not very nice

```xml
<mo:MusicArtist>
  rdf:about="http://musicbrainz.org/artist/b10bbbfccf9e-42e0-be17-e2c3e1d260d#_">
  <foaf:name>The Beatles</foaf:name>
  <owl:sameAs rdf:resource="http://dbpedia.org/resource/The_Beatles"/>
</mo:MusicArtist>
```
Describing Data

Vocabularies

• Collections of defined **properties** and **classes** of resources.
  • Classes group together similar resources.
    • `mo:MusicArtist`, `foaf:Person`, `geo:Point`
  • Properties denote relationships
    • `mo:member`, `foaf:name`, `rdf:type`

• Terms from well-known vocabularies should be **reused** wherever possible.

• New terms should only be defined when you cannot find required terms in existing vocabularies.
## Describing Data

### Vocabularies

A set of well-known vocabularies has evolved in the Semantic Web community. *Some* of them are:

<table>
<thead>
<tr>
<th>Vocabulary</th>
<th>Description</th>
<th>Classes and Relationships</th>
</tr>
</thead>
<tbody>
<tr>
<td>Friend-of-a-Friend <em>(FOAF)</em></td>
<td>Vocabulary for describing people.</td>
<td>foaf:Person, foaf:Agent, foaf:name,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>foaf:knows, foaf:member</td>
</tr>
<tr>
<td>Dublin Core <em>(DC)</em></td>
<td>Defines general metadata attributes.</td>
<td>dc:FileFormat, dc:MediaType,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dc:creator, dc:description</td>
</tr>
<tr>
<td>Semantically-Interlinked Online Communities <em>(SIOC)</em></td>
<td>Vocabulary for representing online communities.</td>
<td>sioc:Community, sioc:Forum,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sioc:Post, sioc:follows, sioc:topic</td>
</tr>
<tr>
<td>Music Ontology <em>(MO)</em></td>
<td>Provides terms for describing artists, albums and tracks.</td>
<td>mo:MusicArtist, mo:MusicGroup,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mo:Signal, mo:member, mo:record</td>
</tr>
<tr>
<td>Simple Knowledge Organization System <em>(SKOS)</em></td>
<td>Vocabulary for representing taxonomies and loosely structured knowledge.</td>
<td>skos:Concept, skos:inScheme,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>skos:definition, skos:example</td>
</tr>
</tbody>
</table>
Describing Data

Vocabularies

More extensive lists of well-known vocabularies are maintained by:

• **W3C SWEO Linking Open Data community project**
  http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/CommonVocabularies

• **Mondeca: Linked Open Vocabularies**
  http://labs.mondeca.com/dataset/lov

• **Library Linked Data Incubator Group: Vocabularies in the library domain**
  http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025
Describing Data

Defining Vocabularies:
1. Informal human-readable documentation
2. Formal machine-readable documentation

Class: mo:Festival - stable - level 2

Festival - A festival - musical/artistic event lasting several days, like Glastonbury, Rock Am Ring... We might decompose this event (which is in fact just a classification of the space/time region related to a particular festival) using hasSubEvent in several performances at different space/time.

sub-class-of: event:Event
Later on ...

RDF-S – RDF Schema

Semantic Web Stack
Berners-Lee (2006)
Later on ...

OWL – Web Ontology Language

Semantic Web Stack
Berners-Lee (2006)
Later on ...

SPARQL – * Protocol and RDF Query Language

Semantic Web Stack
Berners-Lee (2006)
Acknowledgements

• Alexander Mikroyannidis
• Alice Carpentier
• Andreas Harth
• Andreas Wagner
• Andriy Nikolov
• Barry Norton
• Daniel M. Herzig
• Elena Simperl
• Günter Ladwig
• Inga Shamkhalov
• Jacek Kopecky
• John Domingue

• Juan Sequeda
• Kalina Bontcheva
• Maria Maleshkova
• Maria-Esther Vidal
• Maribel Acosta
• Michael Meier
• Ning Li
• Paul Mulholland
• Peter Haase
• Richard Power
• Steffen Stadtmüller
INTRODUCTION TO: LINKED DATA